| DATASHEET

bluemetrix
the data control company

BDM Schema Evolution

Propagate all schema changes that occur at source, as they happen to the corresponding data store in your Data Lake.

PRODUCT DESCRIPTION

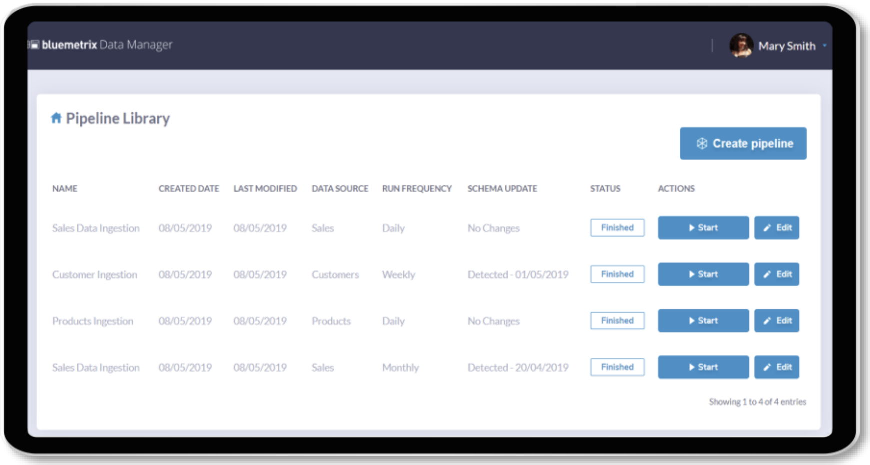
The BDM Schema Evolution module maintains a central record of the schemas of all data sources that are feeding into the Data Lake, along with a versioning system which records all changes to these schemas and when these changes were made. It uses this information to ensure that the corresponding schemas on the data lake are kept consistent with the schemas of the data sources.

BUSINESS CHALLENGE

One of the key business challenges facing data professionals across all industries today is to ensure that all people within the organization are working from the same datasets.

The issue is more complex than just keeping a record of the schemas at the data sources and data destinations and keeping them in sync. It is also necessary to understand how these schema changes affect existing pipelines which are expecting to receive a specific data schema and now have a different schema available to them.

With some data lakes ingesting thousands of different files and tables onto their data lake each week, and running thousands of different pipelines from multiple departments, the only way to guarantee consistency is to automate the process.



Pipeline Library

NAME	CREATED DATE	LAST MODIFIED	DATA SOURCE	RUN FREQUENCY	SCHEMA UPDATE	STATUS	ACTIONS
Sales Data Ingestion	08/05/2019	08/05/2019	Sales	Daily	No Changes	Finished	▶ Start Edit
Customer Ingestion	08/05/2019	08/05/2019	Customers	Weekly	Detected - 01/05/2019	Finished	▶ Start Edit
Products Ingestion	08/05/2019	08/05/2019	Products	Daily	No Changes	Finished	▶ Start Edit
Sales Data Ingestion	08/05/2019	08/05/2019	Sales	Monthly	Detected - 20/04/2019	Finished	▶ Start Edit

Showing 1 to 4 of 4 entries

KEY FEATURES

- Schema changes at source are recognized as they occur
- Each Schema is linked to the individual pipelines that feed off it
- The pipelines owners also receive notifications on the schema changes as they are identified
- Works with many different data sources i.e. databases, files, etc.

KEY BENEFITS

- **Data Consistency**
  Data consistency is guaranteed between the data sources and the data destinations

- **Up to date Pipelines**
  Pipeline owners are always informed of changes to their data sources, allowing them to keep their pipelines up to date

- **Enhanced Audit Trail**
  Versioning of data sources and pipelines provides detailed audit trails for all pipelines run

BDM SOLUTION

We have developed a solution which delivers on these two issues:

Schema Consistency:

- A central repository is kept of all schema's that feed the data lake
- The schema consistency of the data sources is read several times each day (configurable - with the ability to check schemas on-demand) and if changes have occurred these changes are recorded and stored in the central repository
- This repository is the basis for a Versioning system which records all changes to schemas.

Pipeline Consistency:

- All pipelines are recorded in the central repository and a record is created of all schemas that they are consuming data from
- As changes are recorded in these schemas, the owner of the pipeline is informed and given the option to ignore these changes or to upgrade their pipeline
- A version control system is also kept in place for pipelines to record which data source each pipeline is running off

We also guarantee consistency across permissions – by this I mean that all users on the data lake will inherit their permissions from the original data sources. If a schema is changed and a user does not have permission to access/view the changed data, they will not be informed or aware of the schema change in their pipeline versioning.

TECHINCAL REQUIREMENTS

Spark V 1.x or V 2.x
Atlas V 0.8 or above

APPOINTMENT

For more assistance, you can schedule a one-to-one appointment via Skype, Zoom, WebEx or phone.

FOR MORE INFORMATION.

To learn more about BDM and the Schema Evolution module
Please visit www.bluemetrix.com
| Europe +353 21 4212223 | info@bluemetrix.com